FoP: Never-ending face recognition and data lifting

Julien Subercaze, Christophe Gravier

FoP: Never-ending face recognition and data lifting.

Julien Subercaze
Université de Saint-Etienne,
10, rue Tréfilerie
Saint-Étienne, F-42000, France
julien.subercaze@univ-st-etienne.fr

Christophe Gravier
Université de Saint-Etienne
10, rue Tréfilerie
Saint-Étienne, F-42000, France
christophe.gravier@univ-st-etienne.fr

ABSTRACT

In this demonstration, we present `Faces of Politics` (`FoP`), a face detection system from pictures illustrating news articles. The first iteration of the face recognition model propelling `FoP` was trained using Freebase data about politicians and their pictures. `FoP` is a never-ending system: when a new face is recognized, the learned model is updated accordingly. At this step, `FoP` is also giving data in return to the LoD cloud that fed him in the first place: it leverages visual knowledge as Linked Data.

Categories and Subject Descriptors

H.3.5 [**Information Systems**]: Information storage and retrieval—*Online Information Services*; I.2.6 [**Artificial intelligence**]: Learning—*Learning*

Keywords

face recognition, data lifting, never-ending learner

1. INTRODUCTION

The size of the Web makes it difficult to leverage knowledge from all this corpus in a single shot. Learning structured knowledge from the Web is therefore an incremental process. This process never ends, even if it is only because new data is generated daily. This principle was popularized by NELL, the never-ending learner [3], that has been running continuously, and has learnt more than 50 millions of facts by itself. This principle was recently applied at a Web scale for multimedia data by the same group of researchers[4]. In `FoP`, we are interested in a particular sort of multimedia data: the presence of people in the image. In this demonstration we are focusing on images of politicians. Ultimately, `FoP` will be able to continuously learn new representation of people in a picture from people already known in its database.

2. RELATED WORK

With the assumption that video objects are an opaque information nutshell for crawlers, [6] presents a generic crowdsourcing framework for automatic and scalable semantic annotations of HTML5 videos. This framework is easing the leveraging of Linked Data based on the Event Ontology[1], which includes the `Agent` class whose instances can be persons. Another related work is the Flickr wrapper [2]. The Flickr wrapper intends to extend Wikipedia with user-generated semantic annotations on Flickr pictures. However, the linkage possibility are limited to declare that a picture is "related" to a Wikipedia page (related can be pretty vague), and it is not possible to link only a sub-part of the picture (e.g. faces). [5] presented FANS, a face annotation framework. The FANS inner model is first trained using a manually constructed dataset resulting from google queries on 6,025 persons. By applying a Locality-Sensitive Hash function to the natural feature points of detected faces, the authors demonstrated a scalable Web image retrieval engine. However, FANS do not interact with the Web of Data, and the model is not evolving with time.

3. FOP ARCHITECTURE

The `FoP` system data leveraging flow is illustrated in Figure 1. The exchange of data between `FoP` and the LoD-cloud is giving rise to raising edges of data called "`ticks`", while the falling edges (from the LoD-cloud to `FoP`) are respectively called "`tocks`".

3.1 Tocks

`Tocks` are data queried from the LoD-cloud that enrich the `FoP` learner. The first `tock` drived the training of a first face recognition model. For this, the Local Binary Patterns algorithm [1] is a very robust face recognition algorithm that is less sensitive to lighting conditions than other standard algorithms. Moreover it is not computationally expensive to perform an update of the model, which is a unique feature among holistic methods [7]. Face recognition algorithms are however very sensitive to parameters settings. We conducted several tests to determine the best parameter set to use at `tock` phase. The task of face recognition is still a hard research issue : it is not possible to maintain high values for both precision and recall. As our system will handle a large volume of pictures, and that our goal is to provide data with the highest level of correctness, we trade recall for precision. In order to determine the best parameter that

[1] `http://motools.sourceforge.net/event/event.html`

Hypertext 2014 Santiago, Chile

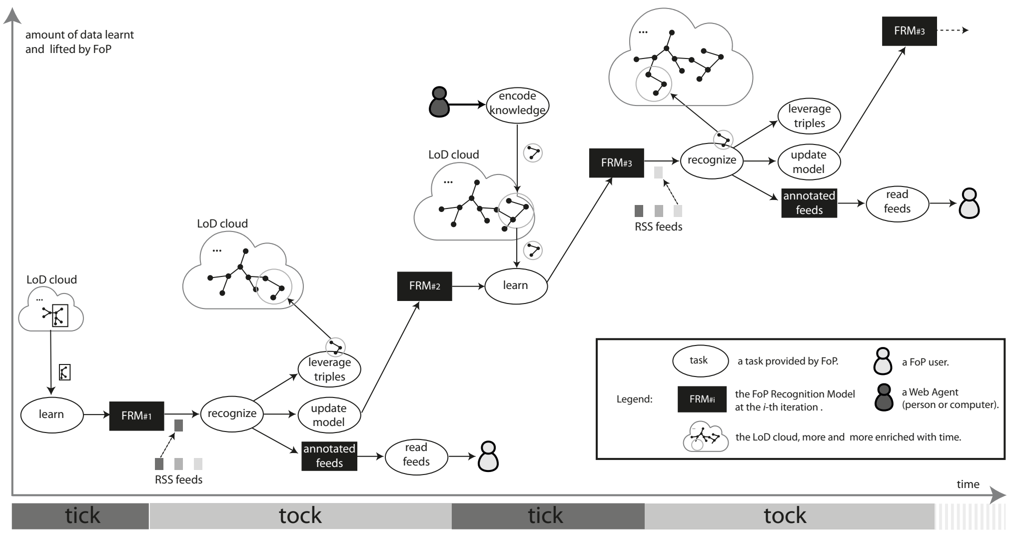

Figure 1: Never-ending face recognizing and semantizing tick/tock model.

maximize precision while preserving recall, we learnt a first model with French politicians in national news. We extract the list of living politicians having government positions in France along to their pictures in Freebase. We then use these pictures and their associated resource to train a first version of the face recognition model. For this first model, we experimentally set the single parameter (which is the maximum authorized distance between two pictures to be considered neighbors) of the Binary Local Pattern algorithm to 57. This value maximizes the recall for a near-perfect precision (0.97). However, the recall still remains low (0.15) and very few pictures would be handled. That is the reason why other `tocks` occur at regular interval of time. At further `tocks`, `FoP` queries again the Freebase database for new Picture–Person links. At each `tock`, the `FoP` face recognition model is therefore enhanced using LoD-based data. `FoP`gives knowledge back to the LoD based data at `tick`.

3.2 Ticks

A `tick` is initiated by receiving an update of one of the RSS feeds to which `FoP` has subscribed. For each article `FoP` extracts pictures, detects faces, and matches them against the previously trained model. If a person is recognized using the face recognition model, we validate its presence by searching the article text for his/her name. If validated, annotated feeds articles are added to the internal database, and the model is updated using the newly detected face. A `tick` can therefore also generate model updates of the `FoP` recognition model. Updating the model allows us to overcome the initial low recall issue. For a threshold of 57, we maintain a very high precision (above .93). In the mean time, recall greatly improved from .15 to .35. For greater values of the threshold the recall improvement is far greater but it implies serious degradation of the precision to unacceptable values, that would result in falsely labeled pictures. Aside from updating the model and storing annotated feeds articles for later consultation by `FoP` users, `FoP` leverages Linked Data at `ticks`. The ontology for media resources (`ma-ont`[2]) defines media fragment as sub-parts of a media. Mediafragments[3] makes it possible to specify rectangular clipping of images by appending its coordinates to the original URI. This is particularly useful for identifying several persons in a picture, since one can specify the region where the face of a person is located. Due to domain coverage shortcomings of `ma-ont`, we defined an object property `IsInPicture` whose domain is a `foaf:Person` and whose range is the intersection of `ma-ont#Image` and `ma-ont#MediaFragment`. This enables to search for pictures containing multiple people. For example one is able to retrieve in one query pictures containing both France's president and prime minister.

4. DEMONSTRATION OVERVIEW

Our `FoP` never ending learner is available online at `http://demo-satin.telecom-st-etienne.fr/facesofpolitics`.

In this demonstration, `FoP` is subscribing to the major French politicians news feeds. We will showcase the following to the conference attendees. First, through the `FoP` user interface we shall demonstrate `FoP` in action by accessing latest feeds with recognized person(s). We will present discovered persons along the part of the multimedia in which they were found, along the context of the original feed article. Second, by switching to another GUI area, we will query `FoP` data using SPARQL. The demonstration will encompass article retrieval based on the presence of one known person in the `FoP` triplestore. Then, we will also demonstrate how it is possible to create a curation of articles from different RSS feeds by specifying several persons that must be present in the article illustration.

Moreover, we also provide a public SPARQL endpoint and a Web interface in order to query the Linked Data leveraged by `FoP` to LOD-cloud resources. Full RDF Dumps and a public SPARQL endpoint are also available at the `FoP`website.

[2] `http://www.w3.org/TR/mediaont-10/`
[3] `http://www.w3.org/TR/media-frags/`

5. REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] C. Becker and C. Bizer. flickr wrappr–precise photo association, 2011.

[3] A. Carlson, J. Betteridge, B. Kisiel, B. Settles, E. R. Hruschka Jr, and T. M. Mitchell. Toward an architecture for never-ending language learning. In *AAAI*, 2010.

[4] X. Chen, A. Shrivastava, and A. Gupta. Neil: Extracting visual knowledge from web data. December 2013.

[5] S. C. H. Hoi, D. Wang, I. Y. Cheng, E. W. Lin, J. Zhu, Y. He, and C. Miao. Fans: face annotation by searching large-scale web facial images. In *WWW (Companion Volume)*, pages 317–320, 2013.

[6] T. Steiner, R. Verborgh, R. Van de Walle, M. Hausenblas, and J. G. Vallé s. Crowdsourcing event detection in youtube video. In M. Van Erp, W. R. Van Hage, L. Hollink, A. Jameson, and R. l. Troncy, editors, *Proceedings of the 1st workshop on detection, representation, and exploitation of events in the semantic web*, pages 58–67, 2011.

[7] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, 35(4):399–458, 2003.